Supplementary Material

A Impact of Generations

In Fig. 5 of the main paper, we demonstrated that MAPMX-P has fast convergence speed in the early generations. One question is whether fitness still improves after 20 generations. In this section, we perform a comparison of train and test R^2 scores between 20 and 100 generations to verify the improvement after 20 generations. The experimental results are presented in Fig. A1 , and the results show that both training and test scores can still significantly improve after 20 generations. Specifically, Fig. A1a shows that increasing the number of generations from 20 to 100 can significantly improve training R^2 on 91 out of 98 datasets. In terms of the test R^2 scores, as shown in Fig. A1b, the test performance is better on 43 datasets when increasing the number of generations from 20 to 100, and worse on only 4 datasets. In general, these results reveal that MAPMX-P can still improve significantly after a fast convergence period in the early generations. Note that the significant increase in training R^2 does not lead to a significant increase in test R^2 on many datasets, which indicates that GP tends to overfit on some datasets. Thus, it is worthwhile to develop an overfitting control technique in the future to further unleash the capability of GSMX operators.

(a) Training R^2 (b) Test R^2

Fig. A1: Summary of Statistical Comparison over 98 Datasets between 20 and 100 Generations